RAIRO-Oper. Res. 55 (2021) S1395–S1410
https://doi.org/10.1051/ro/2020030

RAIRO Operations Research
www.rairo-ro.org

A SPLITTING SUBGRADIENT ALGORITHM FOR SOLVING EQUILIBRIUM PROBLEMS INVOLVING THE SUM OF TWO BIFUNCTIONS AND APPLICATION TO COURNOT-NASH MODEL

Phung Minh Duc[1] and Xuan Thanh Le[2,*]

Abstract. In this paper we propose a splitting subgradient algorithm for solving equilibrium problems involving the sum of two bifunctions. At each iteration of the algorithm, two strongly convex subprograms are required to solve separately, one for each component bifunction. In contrast to the splitting algorithms previously proposed in Anh and Hai (*Numer. Algorithms* **76** (2017) 67–91) and Hai and Vinh (*Rev. R. Acad. Cienc. Exactas Fis. Nat. Ser. A Mat.* **111** (2017) 1051–1069), our algorithm is convergent for paramonotone and strongly pseudomonotone bifunctions without any Lipschitz type as well as Hölder continuity condition of the bifunctions involved. Furthermore, we show that the ergodic sequence defined by the algorithm iterates converges to a solution without paramonotonicity property. Some numerical experiments on differentiated Cournot-Nash models are presented to show the behavior of our proposed algorithm with and without ergodic.

Mathematics Subject Classification. 90C33, 90C56.

Received June 17, 2019. Accepted March 18, 2020.

1. Introduction

Let $\mathcal{H}$ be a real Hilbert space endowed with weak topology defined by the inner product $\langle \cdot, \cdot \rangle$ and its induced norm $\| \cdot \|$. Let $C \subseteq \mathcal{H}$ be a nonempty closed convex subset and $f : \mathcal{H} \times \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ a bifunction such that $f(x, y) < +\infty$ for every $x, y \in C$. The equilibrium problem defined by the Nikaidô–Isoda–Fan inequality that we are going to consider in this paper is given as

$$\text{Find } x \in C : f(x, y) \geq 0 \quad \forall y \in C. \tag{EP}$$

This inequality first was used in 1955 by Nikaidô and Isoda [23] in convex game models. Then in 1972 Ky Fan [9] called this inequality a minimax one and established existence theorems for Problem (EP). After the appearance of the paper by Blum and Oettli [6], Problem (EP) has been attracted much attention of researchers. It has been shown in [4, 6, 19] that some important problems such as optimization, variational inequality, Kakutani fixed point and Nash equilibrium can be formulated in the form of (EP). Many papers concerning the solution

Keywords. Monotone equilibria, splitting algorithm, ergodic sequence.

[1] Ho Chi Minh University of Natural Resources and Environment, Ho Chi Minh City, Vietnam.

[2] Institute of Mathematics, Vietnam Academy of Science and Technology, Hanoi, Vietnam.

*Corresponding author: lxthanh@math.ac.vn

Article published by EDP Sciences

existence, stabilities as well as algorithms for Problem (EP) have been published (see *e.g.*, [11,13,16,20,25–27], the survey paper [4], and the interesting monograph [5]).

A basic method for Problem (EP) is the subgradient (or projection) one, where the sequence of iterates is defined by taking

$$x^{k+1} = \min\left\{\lambda_k f\left(x^k, y\right) + \frac{1}{2}\|y - x^k\|^2 : y \in C\right\}, \tag{1.1}$$

with λ_k is some appropriately chosen real number. Note that in the variational inequality case, where $f(x,y) := \langle F(x), y - x\rangle$, the iterate x^{k+1} defined by (1.1) becomes

$$x^{k+1} = P_C\left(x^k - \lambda_k F\left(x^k\right)\right),$$

where P_C stands for the metric projection onto C. It is well known that under certain conditions on the parameter λ_k, the projection method is convergent if f is strongly pseudomonotone or paramonotone [7,13]. However when f is monotone, it may fail to converge. In order to obtain convergent algorithms for monotone, even pseudomonotone, equilibrium problems, the extragradient method first proposed by Korpelevich [15] for the saddle point and related problems has been extended to equilibrium problems [25]. In this extragradient algorithm, at each iteration, it requires solving the two strongly convex programs

$$y^k = \min\left\{\lambda_k f\left(x^k, y\right) + \frac{1}{2}\|y - x^k\|^2 : y \in C\right\}, \tag{1.2}$$

$$x^{k+1} = \min\left\{\lambda_k f\left(x^k, y\right) + \frac{1}{2}\|y - y^k\|^2 : y \in C\right\}, \tag{1.3}$$

which may cause computational cost. In order to reduce the computational cost, several convergent algorithms that require solving only one strongly convex program or computing only one projection at each iteration have been proposed for some classes of bifunctions such as strongly pseudomonotone and paramonotone with or without using an ergodic sequence (see *e.g.*, [2,7,27]). In another direction, also for the sake of reducing computational cost, some splitting algorithms have been developed (see *e.g.*, [1,10,18]) for monotone equilibrium problems where the bifunction f can be decomposed into the sum of two bifunctions. In these algorithms the convex subprograms (resp. regularized subproblems) involving the bifunction f can be replaced by two convex subprograms (resp. regularized subproblems), one for each component bifunction independently. However, for the convergence, these algorithms require Lipschitz type or Hölder continuity conditions on the involved bifunctions.

For solving the equilibrium problems, in this paper we propose a splitting subgradient algorithm with the following main features. At each iteration, it requires solving only one strongly convex program. Similar to the algorithm in [1,10], in the case where the bifunction f can be represented as the sum of two bifunctions $f_1 + f_2$, this strongly convex subprogram can be replaced by two strongly convex subprograms, one for each component bifunction f_1 and f_2. Nevertheless, for the convergence, our algorithm does not require any additional conditions such as Hölder continuity and Lipschitz type condition of these bifunctions. Furthermore, we show that the ergodic sequence defined by the iterates obtained by our algorithm is convergent to a solution without paramonotonicity property. We apply the two versions of the algorithm (with and without ergodic sequence) for solving some versions of a differentiated Cournot-Nash model. Some preliminary computational results for comparing the proposed algorithms with the ones in [2,27] are reported.

The remaining part of the paper is organized as follows. The next section gives preliminaries containing some lemmas that will be used in proving the convergence of the proposed algorithm. Section 3 is devoted to the description of the algorithm and its convergence analysis for both versions with and without ergodic. Some numerical experiments are presented in Section 4. Section 5 closes this paper with some conclusions.

2. Preliminaries

We recall from [4] the following well-known definition on monotonicity of bifunctions.

Definition 2.1. A bifunction $f : \mathcal{H} \times \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is said to be

(i) strongly monotone on C with modulus $\beta > 0$ (shortly β-strongly monotone) if

$$f(x,y) + f(y,x) \leq -\beta\|y-x\|^2 \quad \forall x, y \in C;$$

(ii) monotone on C if

$$f(x,y) + f(y,x) \leq 0 \quad \forall x, y \in C;$$

(iii) strongly pseudomonotone on C with modulus $\beta > 0$ (shortly β-strongly pseudomonotone) if

$$f(x,y) \geq 0 \implies f(y,x) \leq -\beta\|y-x\|^2 \quad \forall x, y \in C;$$

(iv) pseudomonotone on C if

$$f(x,y) \geq 0 \implies f(y,x) \leq 0 \quad \forall x, y \in C.$$

(v) paramonotone on C with respect to a set S if

$$x^* \in S, x \in C \text{ and } f(x^*, x) = f(x, x^*) = 0 \text{ implies } x \in S.$$

Clearly in the case of optimization problem when $f(x,y) = \varphi(y) - \varphi(x)$, the bifunction f is paramonotone on C with respect to the solution set of the problem $\min_{x \in C} \varphi(x)$. It is obvious that (i) $\implies$ (ii) $\implies$ (iv) and (i) $\implies$ (iii) $\implies$ (iv). Note that a strongly pseudomonotone bifunction may not be monotone. Paramonotone bifunctions have been used in *e.g.*, [27, 28]. Some properties of paramonotone operators can be found in [12], where a multivalued monotone operator T is called paramonotone on $C \subseteq \mathrm{dom}T$ if for every $x, y \in C$ we have $u \in T(x)$, $v \in T(y)$, $\langle u - v, x - y\rangle = 0$ implies $u \in T(y)$, $v \in T(x)$. The following remark gives a connection between paramonotone operators and paramonotone bifunctions.

Remark 2.2. Let T be a convex, compact valued multivalued operator on C. The bifunction $f(x,y) := \max_{u \in T(x)} \langle u, y - x\rangle$ is paramonotone with respect to the solution set $S(C,f)$ of Problem (EP) if T is paramonotone on C.

Proof. Suppose T is paramonotone on C and let $x^* \in S(C,f)$, $\bar{x} \in C$ such that $f(x^*, \bar{x}) = f(\bar{x}, x^*) = 0$. On one hand, since $x^* \in S(C,f)$, there exists $u^* \in T(x^*)$ such that $\langle u^*, x - x^*\rangle \geq 0$ for every $x \in C$. In particular we have $\langle u^*, \bar{x} - x^*\rangle \geq 0$. On the other hand, by definition we have $f(x^*, \bar{x}) = \max\{\langle u, \bar{x} - x^*\rangle : u \in T(x^*)\}$. Since $f(x^*, \bar{x}) = 0$, this means $\langle u^*, \bar{x} - x^*\rangle \leq 0$. So we obtain

$$\langle u^*, \bar{x} - x^*\rangle = 0. \tag{2.1}$$

Since $f(\bar{x}, x^*) = 0$, there exists $\bar{u} \in T(\bar{x})$ satisfying $\langle \bar{u}, x^* - \bar{x}\rangle = 0$, which together with (2.1) implies $\langle u^* - \bar{u}, x^* - \bar{x}\rangle = 0$. By paramonotonicity of T, we have $u^* \in T(\bar{x})$, which implies

$$f(\bar{x}, x) := \max_{u \in T(\bar{x})} \langle u, x - \bar{x}\rangle \geq \langle u^*, x - \bar{x}\rangle = \langle u^*, x - x^*\rangle + \langle u^*, x^* - \bar{x}\rangle \geq 0 \quad \forall x \in C.$$

Hence $\bar{x}$ solves the problem (EP), *i.e.*, $\bar{x} \in S(C,f)$. Therefore f is paramonotone with respect to $S(C,f)$. □

The following well known lemmas will be used for proving the convergence of the algorithm to be described in the next section.

Lemma 2.3 (see [29], Lem. 1)**.** *Let $\{\alpha_k\}$ and $\{\sigma_k\}$ be two sequences of nonnegative numbers such that $\alpha_{k+1} \leq \alpha_k + \sigma_k$ for all $k \in \mathbb{N}$, where $\sum_{k=1}^{\infty} \sigma_k < \infty$. Then the sequence $\{\alpha_k\}$ is convergent.*

Lemma 2.4 (see [3], Lem. 2.39)**.** *Let $\mathcal{H}$ be a Hilbert space, $\{x^k\}$ a sequence in $\mathcal{H}$ and C be a nonempty subset of $\mathcal{H}$. Suppose that:*

(i) *For every $x \in C$, $\left\{\|x^k - x\|\right\}_{k\in\mathbb{N}}$ converges.*
(ii) *Every weak cluster point of the sequence $\left\{x^k\right\}$ belongs to C.*

Then the sequence $\left\{x^k\right\}$ converges weakly to a point in C.

Lemma 2.5 (see [24])**.** *Let $\mathcal{H}$ be a Hilbert space, $\left\{x^k\right\}$ a sequence in $\mathcal{H}$. Let $\{r_k\}$ be a sequence of nonnegative number such that $\sum_{k=1}^{\infty} r_k = +\infty$ and set $z^k := \dfrac{\sum_{i=1}^k r_i x^i}{\sum_{i=1}^k r_i}\cdot$ Assume that there exists a nonempty, closed convex set $S \subset \mathcal{H}$ satisfying:*

(i) *For every $z \in S$, $\lim_{n\to\infty} \|z^k - z\|$ exists.*
(ii) *Any weakly cluster point of the sequence $\left\{z^k\right\}$ belongs to S.*

Then the sequence $\left\{z^k\right\}$ converges weakly to a point in S.

Lemma 2.6 (see [31])**.** *Let $\{\lambda_k\}, \{\delta_k\}, \{\sigma_k\}$ be sequences of real numbers such that*

(i) $\lambda_k \in (0,1)$ *for all* $k \in \mathbb{N}$*;*
(ii) $\sum_{k=1}^{\infty} \lambda_k = +\infty$*;*
(iii) $\limsup_{k\to+\infty} \delta_k \le 0$*;*
(iv) $\sum_{k=1}^{\infty} |\sigma_k| < +\infty$.

Suppose that $\{\alpha_k\}$ is a sequence of nonnegative real numbers satisfying

$$\alpha_{k+1} \le (1-\lambda_k)\alpha_k + \lambda_k\delta_k + \sigma_k \quad \forall k \in \mathbb{N}.$$

Then we have $\lim_{k\to+\infty} \alpha_k = 0$.

3. The algorithm and its convergence

In what follows, for the following equilibrium problem

$$\text{find } x \in C : f(x,y) \ge 0 \quad \forall y \in C \tag{EP}$$

we suppose that $f(x,y) = f_1(x,y) + f_2(x,y)$ and that $f_i(x,x) = 0$ $(i = 1,2)$ for every $x, y \in C$. The following assumptions for the bifunctions f, f_1, f_2 will be used in the sequel.

(A1) For each $i = 1,2$ and each $x \in C$, the function $f_i(x,\cdot)$ is convex and sub-differentiable, while for each $y \in C$ the function $f(\cdot,y)$ is weakly upper semicontinuous on C.
(A2) If $\left\{x^k\right\} \subset C$ is bounded, then for each $i = 1,2$, the sequence $\left\{g_i^k\right\}$ with $g_i^k \in \partial_2 f_i\left(x^k, x^k\right)$ is bounded.
(A3) The bifunction f is monotone on C.

Assumption (A2) has been used in *e.g.*, [28]. Note that Assumption (A2) is satisfied if the functions f_1 and f_2 are jointly weakly continuous on an open convex set containing C (see [30], Prop. 4.1).

The dual problem of (EP) is

$$\text{find } x \in C : f(y,x) \le 0 \quad \forall y \in C. \tag{DEP}$$

We denote the solution sets of (EP) and (DEP) by $S(C,f)$ and $S^d(C,f)$, respectively. A relationship between $S(C,f)$ and $S^d(C,f)$ is given in the following lemma.

Lemma 3.1 (see [14], Prop. 2.1)**.**

(i) *If $f(\cdot,y)$ is weakly upper semicontinuous and $f(x,\cdot)$ is convex for all $x, y \in C$, then $S^d(C,f) \subseteq S(C,f)$.*
(ii) *If f is pseudomonotone, then $S(C,f) \subseteq S^d(C,f)$.*

Therefore, under the assumptions (A1)–(A3) one has $S(C,f) = S^d(C,f)$. In this paper we suppose that $S(C,f)$ is nonempty. The algorithm below is a subgradient one for paramonotone or strongly pseudomonotone equilibrium problems (EP). The stepsize is taken as in the subgradient method for nonsmooth convex optimization problems.

Algorithm 1. A splitting subgradient algorithm for solving paramonotone or strongly pseudomonotone equilibrium problems.

Initialization: Seek $x^0 \in C$. Choose a sequence $\{\beta_k\}_{k\geq 0} \subset \mathbb{R}$ satisfying the following conditions

$$\sum_{k=0}^{\infty} \beta_k = +\infty, \quad \sum_{k=0}^{\infty} \beta_k^2 < +\infty.$$

Iteration $k = 0, 1, \ldots$:

Take $g_1^k \in \partial_2 f_1\left(x^k, x^k\right), g_2^k \in \partial_2 f_2\left(x^k, x^k\right)$.

Compute

$$\eta_k := \max\left\{\beta_k, \|g_1^k\|, \|g_2^k\|\right\},\ \lambda_k := \frac{\beta_k}{\eta_k},$$
$$y^k := \arg\min\left\{\lambda_k f_1\left(x^k, y\right) + \frac{1}{2}\|y - x^k\|^2 \mid y \in C\right\},$$
$$x^{k+1} := \arg\min\left\{\lambda_k f_2\left(x^k, y\right) + \frac{1}{2}\|y - y^k\|^2 \mid y \in C\right\}.$$

Theorem 3.2. *In addition to the assumptions (A1)–(A3) we suppose that f is paramonotone on C, and that either int $C \neq \emptyset$ or for each $x \in C$ both bifunctions $f_1(x,\cdot)$, $f_2(x,\cdot)$ are continuous at a point in C. Then the sequence $\left\{x^k\right\}$ generated by Algorithm 1 converges weakly to a solution of (EP). Moreover, if f is strongly pseudomonotone, then $\left\{x^k\right\}$ strongly converges to the unique solution of (EP).*

Proof. (i) We first show that, for each $x^* \in S(f,C)$, the sequence $\left\{\|x^k - x^*\|\right\}$ is convergent.

Indeed, for each $k \geq 0$, for simplicity of notation, let

$$h_1^k(x) := \lambda_k f_1\left(x^k, x\right) + \frac{1}{2}\|x - x^k\|^2,$$
$$h_2^k(x) := \lambda_k f_2\left(x^k, x\right) + \frac{1}{2}\|x - y^k\|^2.$$

Since $f_1\left(x^k,\cdot\right)$ is convex and subdifferentiable by Assumption (A1), the functions h_1^k is strongly convex with modulus 1 and subdifferentiable, which implies that for any $u_1^k \in \partial h_1^k\left(y^k\right)$ we have

$$h_1^k\left(y^k\right) + \left\langle u_1^k, x - y^k\right\rangle + \frac{1}{2}\|x - y^k\|^2 \leq h_1^k(x) \quad \forall x \in C. \tag{3.1}$$

On the other hand, since y^k is a minimizer of $h_1^k(x)$ over C, by the regularity condition and the optimality condition for convex programming, we have $0 \in \partial h_1^k\left(y^k\right) + N_C\left(y^k\right)$ in which $N_C\left(y^k\right)$ is the normal cone of C at y^k. This implies that there exists $u_1^k \in \partial h_1^k\left(y^k\right)$ satisfying $u_1^k \in -N_C\left(y^k\right)$, or equivalently, $\left\langle u_1^k, x - y^k\right\rangle \geq 0$ for all $x \in C$. Hence, from (3.1), for each $x \in C$, it follows that

$$h_1^k\left(y^k\right) + \frac{1}{2}\|x - y^k\|^2 \leq h_1^k(x),$$

i.e.,

$$\lambda_k f_1\left(x^k, y^k\right) + \frac{1}{2}\|y^k - x^k\|^2 + \frac{1}{2}\|x - y^k\|^2 \le \lambda_k f_1\left(x^k, x\right) + \frac{1}{2}\|x - x^k\|^2,$$

or equivalently,

$$\|y^k - x\|^2 \le \|x^k - x\|^2 + 2\lambda_k\left(f_1\left(x^k, x\right) - f_1\left(x^k, y^k\right)\right) - \|y^k - x^k\|^2. \tag{3.2}$$

Using the same argument for x^{k+1}, we obtain

$$\|x^{k+1} - x\|^2 \le \|y^k - x\|^2 + 2\lambda_k\left(f_2\left(x^k, x\right) - f_2\left(x^k, x^{k+1}\right)\right) - \|x^{k+1} - y^k\|^2. \tag{3.3}$$

Combining (3.2) and (3.3) yields

$$\begin{aligned}\|x^{k+1} - x\|^2 &\le \|x^k - x\|^2 - \|y^k - x^k\|^2 - \|x^{k+1} - y^k\|^2\\ &\quad + 2\lambda_k\left(f_1\left(x^k, x\right) + f_2\left(x^k, x\right)\right) - 2\lambda_k\left(f_1\left(x^k, y^k\right) + f_2\left(x^k, x^{k+1}\right)\right)\\ &= \|x^k - x\|^2 - \|y^k - x^k\|^2 - \|x^{k+1} - y^k\|^2\\ &\quad + 2\lambda_k f\left(x^k, x\right) - 2\lambda_k\left(f_1\left(x^k, y^k\right) + f_2\left(x^k, x^{k+1}\right)\right).\end{aligned} \tag{3.4}$$

From $g_1^k \in \partial_2 f_1\left(x^k, x^k\right)$ and $f_1\left(x^k, x^k\right) = 0$, it follows that

$$f_1\left(x^k, y^k\right) - f_1\left(x^k, x^k\right) \ge \left\langle g_1^k, y^k - x^k\right\rangle,$$

which implies

$$-2\lambda_k f_1\left(x^k, y^k\right) \le -2\lambda_k\left\langle g_1^k, y^k - x^k\right\rangle. \tag{3.5}$$

By using the Cauchy–Schwarz inequality and the fact that $\|g_1^k\| \le \eta_k$, from (3.5) one can write

$$-2\lambda_k f_1\left(x^k, y^k\right) \le 2\frac{\beta_k}{\eta_k}\eta_k\|y^k - x^k\| = 2\beta_k\|y^k - x^k\|. \tag{3.6}$$

By the same argument, we obtain

$$-2\lambda_k f_2\left(x^k, x^{k+1}\right) \le 2\beta_k\|x^{k+1} - x^k\|. \tag{3.7}$$

Replacing (3.6) and (3.7) to (3.4) we get

$$\begin{aligned}\|x^{k+1} - x\|^2 &\le \|x^k - x\|^2 + 2\lambda_k f\left(x^k, x\right)\\ &\quad + 2\beta_k\|y^k - x^k\| + 2\beta_k\|x^{k+1} - x^k\| - \|y^k - x^k\|^2 - \|x^{k+1} - y^k\|^2.\end{aligned} \tag{3.8}$$

Take $x = x^k$ in (3.8), since $f\left(x^k, x^k\right) = 0$, we obtain

$$\|x^{k+1} - x^k\|^2 \le 2\beta_k\|y^k - x^k\| + 2\beta_k\|x^{k+1} - x^k\| - \|y^k - x^k\|^2 - \|x^{k+1} - y^k\|^2.$$

It follows that

$$\left(\|x^{k+1} - x^k\| - \beta_k\right)^2 + \left(\|y^k - x^k\| - \beta_k\right)^2 + \|x^{k+1} - y^k\|^2 \le 2\beta_k^2.$$

Hence $\left(\|x^{k+1} - x^k\| - \beta_k\right)^2 \le 2\beta_k^2$, and consequently we have

$$\|x^{k+1} - x^k\| \le 3\beta_k. \tag{3.9}$$

Replacing (3.9) to (3.8), we obtain

$$\begin{aligned}\|x^{k+1} - x\|^2 &\le \|x^k - x\|^2 + 2\lambda_k f\left(x^k, x\right) + 2\beta_k\|y^k - x^k\| + 6\beta_k^2 - \|y^k - x^k\|^2 - \|x^{k+1} - y^k\|^2\\ &\le \|x^k - x\|^2 + 2\lambda_k f\left(x^k, x\right) + 6\beta_k^2 + 2\beta_k\|y^k - x^k\| - \|y^k - x^k\|^2\\ &\le \|x^k - x\|^2 + 2\lambda_k f\left(x^k, x\right) + 7\beta_k^2 - \left(\|y^k - x^k\| - \beta_k\right)^2\\ &\le \|x^k - x\|^2 + 2\lambda_k f\left(x^k, x\right) + 7\beta_k^2.\end{aligned} \tag{3.10}$$

Note that by definition of $x^* \in S(f,C) = S^d(f,C)$ we have $f\left(x^k, x^*\right) \leq 0$. Therefore, by taking $x = x^*$ in (3.10) we obtain

$$\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 + 7\beta_k^2 + 2\lambda_k f\left(x^k, x^*\right) \leq \|x^k - x^*\|^2 + 7\beta_k^2. \tag{3.11}$$

Since $\sum_{k=0}^{\infty} \beta_k^2 < +\infty$ by assumption, in virtue of Lemma 2.3, it follows from (3.11) that the sequence $\left\{\|x^k - x^*\|\right\}$ is convergent.

(ii) We now prove that any cluster point of the sequence $\left\{x^k\right\}$ is a solution of (EP).

Indeed, from (3.11) we have

$$-2\lambda_k f\left(x^k, x^*\right) \leq \|x^k - x^*\|^2 - \|x^{k+1} - x^*\|^2 + 7\beta_k^2 \quad \forall k \in \mathbb{N}.$$

By summing up we obtain

$$2\sum_{k=0}^{\infty} \lambda_k \left(-f\left(x^k, x^*\right)\right) \leq \|x^0 - x^*\|^2 + 7\sum_{k=0}^{\infty} \beta_k^2 < \infty. \tag{3.12}$$

On the other hand, by Assumption (A2) the sequences $\left\{g_1^k\right\}, \left\{g_2^k\right\}$ are bounded. This fact, together with the construction of $\{\beta_k\}$, implies that there exists $M > 0$ such that $\|g_1^k\| \leq M, \|g_2^k\| \leq M, \beta_k \leq M$ for all $k \in \mathbb{N}$, and consequently

$$\eta_k = \max\left\{\beta_k, \|g_1^k\|, \|g_2^k\|\right\} \leq M \quad \forall k \in \mathbb{N}.$$

So we have

$$\sum_{k=0}^{\infty} \lambda_k = \sum_{k=0}^{\infty} \frac{\beta_k}{\eta_k} \geq \frac{1}{M} \sum_{k=0}^{\infty} \beta_k,$$

and since $\sum_{k=0}^{\infty} \beta_k = +\infty$, we obtain

$$\sum_{k=0}^{\infty} \lambda_k = +\infty.$$

Since $f\left(x^k, x^*\right) \leq 0$ for all $k \in \mathbb{N}$, it follows from (3.12) that

$$\limsup f\left(x^k, x^*\right) = 0 \quad \forall x^* \in S(C,f).$$

Fixed $x^* \in S(C,f)$ and let $\left\{x^{k_j}\right\}$ be a subsequence of $\left\{x^k\right\}$ such that

$$\limsup f\left(x^k, x^*\right) = \lim_j f\left(x^{k_j}, x^*\right) = 0.$$

Since $\left\{x^{k_j}\right\}$ is bounded, we may assume that $\left\{x^{k_j}\right\}$ weakly converges to some $\bar{x}$. Since $f\left(\cdot, x^*\right)$ is weakly upper semicontinuous by Assumption (A1), we have

$$f\left(\bar{x}, x^*\right) \geq \lim f\left(x^{k_j}, x^*\right) = 0. \tag{3.13}$$

Then it follows from the monotonicity of f that $f\left(x^*, \bar{x}\right) \leq 0$. On the other hand, since $x^* \in S(C,f)$, by definition we have $f\left(x^*, \bar{x}\right) \geq 0$. Therefore we obtain $f\left(x^*, \bar{x}\right) = 0$. Again, the monotonicity of f implies $f\left(\bar{x}, x^*\right) \leq 0$, and therefore, by (3.13) one has $f\left(\bar{x}, x^*\right) = 0$. Since $f\left(x^*, \bar{x}\right) = 0$ and $f\left(\bar{x}, x^*\right) = 0$, it follows from paramonotonicity of f that $\bar{x}$ is a solution to (EP).

Thus it follows from (i), (ii), and Lemma 2.4 that the sequence $\left\{x^k\right\}$ converges weakly to a solution to (EP).

We now turn to the case that f is strongly pseudo-monotone. By this assumption, (EP) has a unique solution (see [21], Prop. 1). Let x^* be the unique solution of (EP). By definition of x^* we have

$$f\left(x^*, x\right) \geq 0 \quad \forall x \in C,$$

which, by strong pseudo-monotonicity of f, implies

$$f\left(x, x^*\right) \leq -\beta\|x - x^*\|^2 \quad \forall x \in C. \tag{3.14}$$

By choosing $x = x^k$ in (3.14) and then applying to (3.8) we obtain

$$\|x^{k+1} - x^*\|^2 \leq (1 - 2\beta\lambda_k)\|x^k - x^*\|^2 + 7\beta_k^2 \quad \forall k \in \mathbb{N},$$

which together with the construction of β_k and λ_k, by virtue of Lemma 2.6 with $\delta_k \equiv 0$, implies that

$$\lim_{k\to+\infty} \|x^k - x^*\|^2 = 0,$$

i.e., x^k strongly converges to the unique solution x^* of (EP). □

The following simple example shows that without paramonotonicity, the algorithm may not be convergent.

Example 3.3. Let us consider the following instance of (EP), taken from [8], where $C := \mathbb{R}^2$ and $f(x,y) := \langle Ax, y - x\rangle$ with

$$A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

For all $x, y \in C$ we have $f(x,y) + f(y,x) = \langle A(x-y), y-x\rangle = (x_2 - y_2)(y_1 - x_1) - (x_1 - y_1)(y_2 - x_2) = 0$, so f is monotone on C. It has already shown in [8] that $x^* = (0,0)^T$ is the unique solution of this problem. Note that for all $x = (x_1, x_2)^T \in C$ we have

$$f\left(x^*, x\right) = \left\langle \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right\rangle = 0$$

and

$$f\left(x, x^*\right) = \left\langle \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \begin{bmatrix} -x_1 \\ -x_2 \end{bmatrix} \right\rangle = -x_2x_1 + (-x_1)(-x_2) = 0.$$

Hence $f\left(x^*, x\right) = f\left(x, x^*\right) = 0$ for all $x \in C$, and therefore f is not paramonotone.

Let $f_1(x,y) = 0$ and $f_2(x,y) = f(x,y) = x_2y_1 - x_1y_2$, then we have $f(x,y) = f_1(x,y) + f_2(x,y)$. Furthermore, f_1, f_2 satisfy assumptions (A1) and (A2). Let $\{\beta_k\}$ be any sequence satisfying the conditions in the initialization step of Algorithm 1. Applying this algorithm with $f = f_1 + f_2$, in iteration k we obtain

$$\begin{aligned} y^k &= \arg\min\left\{\lambda_k f_1\left(x^k, y\right) + \frac{1}{2}\|y - x^k\|^2 \mid y \in \mathbb{R}^2\right\} \\ &= \arg\min\left\{\frac{1}{2}\|y - x^k\|^2 \mid y \in \mathbb{R}^2\right\} \\ &= x^k. \end{aligned}$$

Therefore we have

$$\begin{aligned} \lambda_k f_2\left(x^k, y\right) + \frac{1}{2}\|y - y^k\|^2 &= \lambda_k f_2\left(x^k, y\right) + \frac{1}{2}\|y - x^k\|^2 \\ &= \lambda_k\left(x_2^k y_1 - x_1^k y_2\right) + \frac{1}{2}\left(y_1 - x_1^k\right)^2 + \frac{1}{2}\left(y_2 - x_2^k\right)^2 \\ &= \frac{1}{2}\left(y_1^2 - 2a_1y_1 + \left(x_1^k\right)^2\right) + \frac{1}{2}\left(y_2^2 - 2a_2y_2 + \left(x_2^k\right)^2\right) \\ &= \frac{1}{2}\left(\left(y_1 - a_1\right)^2 + a_1^2 - \left(x_1^k\right)^2\right) + \frac{1}{2}\left(\left(y_2 - a_2\right)^2 + a_2^2 - \left(x_2^k\right)^2\right). \end{aligned}$$

Here $a_1 = x_1^k - \lambda_k x_2^k$ and $a_2 = x_2^k + \lambda_k x_1^k$. It follows that

$$\begin{aligned} x^{k+1} &= \arg\min \left\{ \lambda_k f_2\left(x^k, y\right) + \frac{1}{2}\|y - y^k\|^2 \mid y \in \mathbb{R}^2 \right\} \\ &= \arg\min \left\{ \frac{1}{2}\left(y_1 - a_1\right)^2 + \frac{1}{2}\left(y_2 - a_2\right)^2 \ \mid y \in \mathbb{R}^2 \right\} \\ &= \left(a_1, a_2\right)^T = \left(x_1^k - \lambda_k x_2^k, x_2^k + \lambda_k x_1^k\right)^T. \end{aligned}$$

Thus, $\|x^{k+1}\|^2 = (1+\lambda_k^2)\|x^k\|^2 > \|x^k\|^2$ if $x^k \neq 0$, which implies that the sequence $\left\{x^k\right\}$ does not converge to the solution $x^* = 0$ for any starting point $x^0 \neq 0$. □

In order to obtain the convergence without paramonotonicity we use the iterate x^k to define an ergodic sequence by taking

$$z^k := \frac{\sum_{i=0}^k \lambda_i x^i}{\sum_{i=0}^k \lambda_i}\cdot$$

We then have the following convergence result.

Theorem 3.4. *Under the assumptions (A1)–(A3), the ergodic sequence $\left\{z^k\right\}$ converges weakly to a solution of (EP).*

Proof. In the proof of Theorem 3.2, we have shown that the sequence $\left\{\|x^k - x^*\|\right\}$ is convergent. From the definition of z^k and the Silverman–Toeplitz theorem (see *e.g.*, [22], Thm. 1.1.), the sequence $\left\{\|z^k - x^*\|\right\}$ is convergent, too. In order to apply Lemma 2.4, now we show that all weakly cluster points of $\left\{z^k\right\}$ belong to $S(f, C)$. Indeed, using the inequality (3.11), by taking the sum of its two sides over all indices we have

$$\begin{aligned} 2\sum_{i=0}^k \lambda_i f(x, x^i) &\le \sum_{i=0}^k \left(\|x^i - x\|^2 - \|x^{i+1} - x\|^2 + 7\beta_i^2\right) \\ &= \|x^0 - x\|^2 - \|x^{k+1} - x\|^2 + 7\sum_{i=0}^k \beta_i^2 \\ &\le \|x^0 - x\|^2 + 7\sum_{i=0}^k \beta_i^2. \end{aligned}$$

By using this inequality, from definition of z^k and convexity of $f(x, \cdot)$, we can write

$$\begin{aligned} f\left(x, z^k\right) &= f\left(x, \frac{\sum_{i=0}^k \lambda_i x^i}{\sum_{i=0}^k \lambda_i}\right) \\ &\le \frac{\sum_{i=0}^k \lambda_i f(x, x^i)}{\sum_{i=0}^k \lambda_i} \\ &\le \frac{\|x^0 - x\|^2 + 7\sum_{i=0}^k \beta_i^2}{2\sum_{i=0}^k \lambda_i}\cdot \end{aligned} \tag{3.15}$$

As we have shown in the proof of Theorem 3.2 that

$$\lambda_k = \frac{\beta_k}{\eta_k} \ge \frac{\beta_k}{M}\cdot$$

Since $\sum_{k=0}^{\infty} \beta_k = +\infty$, we have $\sum_{k=0}^{\infty} \lambda_k = +\infty$. Then, it follows from (3.15) that

$$\limsup_k f\left(x, z^k\right) \leq 0. \tag{3.16}$$

Let $\bar{z}$ be any weakly cluster of $\left\{z^k\right\}$. Then there exists a subsequence $\left\{z^{k_j}\right\}$ of $\left\{z^k\right\}$ such that $z^{k_j} \rightharpoonup \bar{z}$. Since $f(x, .)$ is lower semicontinuous, it follows from (3.16) that

$$f(x, \bar{z}) \leq 0.$$

Since this inequality hold for arbitrary $x \in C$, it means that $\bar{z} \in S^d(f, C) = S(f, C)$. Thus it follows from Lemma 2.5 that the sequence $\left\{z^k\right\}$ converges weakly to a point $z^* \in S(f, C)$, which is a solution to (EP). □

4. Numerical experiments

This section aims to evaluate the performance of Algorithm 1 on some numerical examples. We also present some experiments on comparing the performance of our proposed algorithm to the exact version of inexact projected subgradient method proposed in [27] and the ergodic algorithm in [2]. We used MATLAB R2016a for implementing the algorithms, and conducted all experiments on a computer with a Core i5 processor, 16 GB of RAM, and Windows 10.

All the tested instances were taken in finite dimensional setting and designed to satisfy the assumptions (A1)–(A3) as well as the convergence conditions of Algorithm 1 in Theorem 3.2 or Theorem 3.4. Since the iterative points generated by the algorithm are proved to be convergent, in each test we terminated our MATLAB program when the number of iterations is large enough (10^4 in our setting) to obtain an approximate solution. For the same purpose and in spirit of (3.9), we also terminated our MATLAB programs when the distance between two consecutive iteration points is small enough (10^{-3} in our setting, *i.e.*, when $\|x^{k+1} - x^k\| \leq 10^{-3}$, or when $\|z^{k+1} - z^k\| \leq 10^{-3}$ for the ergodic sequences).

4.1. Experiment 1

The problem instance in this subsection is designed for a twofold purpose. First, it is to illustrate our motivation of splitting the bifunction involved in (EP). Second, it is to compare the performance Algorithm 1 to the method proposed by Santos and Scheimberg in [27] without using the ergodic strategy, and the algorithm by Anh *et al.* [2] for the ergodic sequence. For that we consider the following differentiated Cournot-Nash model.

There are n companies producing a common homogeneous commodity. For each company $i = 1, \ldots, n$, let $x_i \geq 0$ be the production level (*i.e.*, the amount of commodity to be produced) of company i, and C_i the strategy set of producing of this company. This means that the condition $x_i \in C_i$ must be satisfied for every $i = 1, \ldots, n$, and $C := C_1 \times \ldots \times C_n$ is the set of feasible production levels $x = (x_1, \ldots, x_n)$ of all these companies. The price (per commodity unit) of company i is given by

$$p_i(x) := \alpha - \sum_{k=1}^{n} \tau_{ik} x_k \quad (\text{with } \alpha > 0, \tau_{ik} \geq 0).$$

For each company $i = 1, \ldots, n$, let $c_i(x) = x^t A_i x$ be the cost for producing its production level, in which $A_i = \left(a^i_{jk}\right)_{n \times n}$ is a square matrix with positive entries. The profit of company i is then given by

$$q_i(x) = x_i p_i(x) - c_i(x).$$

Each company i seeks to maximize its profit by choosing the corresponding production level x_i under the presumption that the production of the other companies are parametric input. In this context, a Nash equilibrium point for the model is a point $x^* \in C$ satisfying

$$q_i\left(x^*[x_i]\right) \leq q_i\left(x^*\right) \quad \forall x \in C, i = 1, \ldots, n,$$

where $x^*\left([x_i]\right)$ stands for the vector obtained from x^* by replacing the component x_i^* by x_i. It means that, if some company i leaves its equilibrium strategy while the others keep their equilibrium positions, then the profit of company i does not increase. The problem of finding such an equilibrium point can be formulated in form of (EP) as follows.

$$(\mathrm{EP}\,(\varphi))\ \text{Find } x \in C \text{ such that } \ \varphi(x,y) \geq 0 \ \text{ for all } y \in C,$$

with $C = C_1 \times \ldots \times C_n$ and

$$\begin{aligned}
\varphi(x,y) &:= \sum_{i=1}^{n} \left(q_i(x) - q_i\left(x\left[y_i\right]\right)\right) \\
&= \sum_{i=1}^{n} \left(\sum_{k=1}^{n} \tau_{ik} x_k \right) (y_i - x_i) + \sum_{i=1}^{n} \tau_{ii} y_i (y_i - x_i) - \alpha \sum_{i=1}^{n} (y_i - x_i) \\
&\quad + \sum_{i=1}^{n} \left(a_{ii}^i y_i^2 + y_i \sum_{k \neq i} (a_{ik}^i + a_{ki}^i) x_k - a_{ii}^i x_i^2 - x_i \sum_{k \neq i} (a_{ik}^i + a_{ki}^i) x_k \right) \\
&= \ \varphi_1(x,y) + \varphi_2(x,y).
\end{aligned}$$

Here $\varphi_1(x,y) = \langle Px + Qy - \bar{\alpha}, y - x \rangle$ with

$$P := \begin{bmatrix} \tau_{11} & \tau_{12} & \ldots & \tau_{1n} \\ \tau_{21} & \tau_{22} & \ldots & \tau_{2n} \\ \cdot & \cdot & \ldots & \cdot \\ \tau_{n1} & \tau_{n2} & \ldots & \tau_{nn} \end{bmatrix}, \quad Q := \begin{bmatrix} \tau_{11} & 0 & \ldots & 0 \\ 0 & \tau_{22} & \ldots & 0 \\ \cdot & \cdot & \ldots & \cdot \\ 0 & 0 & \ldots & \tau_{nn} \end{bmatrix}, \quad \bar{\alpha} := (\alpha, \ldots, \alpha)^T,$$

and $\varphi_2(x,y) = \sum_{i=1}^{n} h_i(x, y_i)$ with

$$h_i(x, y_i) = a_{ii}^i y_i^2 + y_i \sum_{k \neq i} \left(a_{ik}^i + a_{ki}^i\right) x_k - a_{ii}^i x_i^2 - x_i \sum_{k \neq i} \left(a_{ik}^i + a_{ki}^i\right) x_k.$$

Note that by taking $f_1(x,y) := \langle (P+Q)x - \bar{\alpha}, y - x \rangle$ and $f(x,y) := f_1(x,y) + \varphi_2(x,y)$, we have

$$\varphi(x,y) = f(x,y) + \langle Q(y-x), y-x \rangle \, .$$

Since Q is positive semidefinite, it follows from Mastroeni (2003) [17] Proposition 2.1 that the solution set of $(\mathrm{EP}\,(\varphi))$ coincides with the solution set of $(\mathrm{EP}\,(f))$. Therefore, in order to obtain an equilibrium point of $(\mathrm{EP}\,(\varphi))$, we can solve the equilibrium problem $(\mathrm{EP}\,(f))$ associated to the bifunction f.

For a numerical instance, we took $n := 8$, $C_i := [10, 50]$ for all $i = 1, \ldots, n$, and $\alpha := 45$. The discount coefficients τ_{ij} were determined by

$$P := (\tau_{ij})_{n \times n} = \frac{1}{100} \begin{bmatrix} 11.75 & 22 & 2 & 33 & 17 & 4 & 40 & 18 \\ 18 & 13.75 & 1 & 22 & 3 & 2 & 33 & 10 \\ 4 & 1 & 1.25 & 1 & 2 & 5 & 4 & 7 \\ 11 & 15 & 0 & 17 & 10 & 0 & 40 & 10 \\ 5 & 1 & 2 & 27 & 9.5 & 4 & 30 & 10 \\ 8 & 2 & 5 & 2 & 4 & 5 & 8 & 14 \\ 22 & 14 & 8 & 25 & 11 & 16 & 26 & 20 \\ 14 & 2 & 4 & 19 & 17 & 8 & 35 & 12 \end{bmatrix},$$

which implies that

$$Q := \frac{1}{100} \mathrm{diag}\,(11.75, 13.75, 1.25, 17, 9.5, 5, 26, 12)\,.$$

The matrices A_i's are randomly generated in such a way that their entries are integers in $[1, 10]$.

Let $S := \frac{1}{2}\Big((P+Q)+(P+Q)^T\Big)$. Then on one hand we have S is symmetric and positive definite, which can be checked by using MATLAB function "chol". It follows that $P+Q$ is also positive definite. An elementary computation shows that

$$f_1(x,y)+f_1(y,x) = -(y-x)^T(P+Q)(y-x) \leq 0 \quad \forall x,y \in C,$$

which implies the monotonicity of f_1, thanks to the positive definiteness of $P+Q$. We furthermore have

$$\begin{aligned} \varphi_2(x,y)+\varphi_2(y,x) &= -\sum_{i\neq j}(y_i-x_i)(y_j-x_j)\left(a^i_{ij}+a^i_{ji}+a^j_{ij}+a^j_{ji}\right) \\ &= -z^TAz \end{aligned}$$

in which $z=(z_1,\ldots,z_n)^T$ with $z_i = y_i - x_i$ for all $i=1,\ldots,n$, and $A=(a_{ij})_{n\times n}$ is a symmetric matrix with

$$a_{ij} = a_{ji} = \begin{cases} \frac{1}{2}\left(a^i_{ij}+a^i_{ji}+a^j_{ij}+a^j_{ji}\right) & \text{if} \quad i\neq j, \\ 0 & \text{if} \quad i=j. \end{cases}$$

By this formula, the ij-entry of matrix A is defined only by the ij-entries and ji-entries of matrices A^i. In this experiment, the matrices A_i are chosen so that the matrix A is positive semidefinite, and consequently

$$\varphi_2(x,y)+\varphi_2(y,x) = -z^TAz \leq 0,$$

which leads to monotonicity of φ_2. So $f = f_1+\varphi_2$ is also monotone, meaning that condition (A3) is also satisfied.

As a remark, it follows from the positivity of a^i_{ii} that $\varphi_2(x,y)$ is convex quadratic with respect to y. By the construction of f_1 and φ_2, the conditions (A1) and (A2) are also satisfied, so all conditions of Theorem 3.2 are fulfilled and therefore we can apply its version without using ergodic sequence.

On the other hand, it can also be checked that $\mathrm{rank}(S) = \mathrm{rank}(P+Q) = 7 < n = 8$. Thus, by Iusem (1998) [12] Proposition 3.2, the bifunction $f = f_1+\varphi_2$ is paramonotone, which ensures the convergence of Algorithm 1 and the projected subgradient method proposed in [27].

Table 1 presents the performance of three methods: original version of Algorithm 1 (denoted OSGA for short), ergodic version of Algorithm 1 (denoted ESGA for short), and exact version of inexact projected subgradient method proposed in [27] (abbreviated by EPSM). For the first two methods, we set $\beta_k = \frac{10^i}{k+1}$ with $i \in \{0,1,2\}$, while the settings for the last method include $\beta_k = \frac{10^i}{k+1}$ with $i \in \{0,1,2\}$ and $\rho_k = 1$ (see [27] for details of the method). In Table 1 we report the performance of these methods with respect to criteria of running time (in seconds, reported in "Time" columns) and the number of iterations (reported in "Iter" columns). Three choices for starting points were taken:

$$x^a = (20,20,\ldots,20), \quad x^b = (30,30,\ldots,30), \quad x^c = (50,50,\ldots,50).$$

It can be observed from Table 1 that the original version of Algorithm 1 (without using ergodic sequence) has better performance than the ergodic version in both criteria: the number of executed iterations and running time. In sense of running time, method EPSM performs best. It is due to the main operation in EPSM is projection onto the set C, which can be quickly done thanks to the box structure of C. In another perspective, the original version of Algorithm 1 performs best in sense of the number of executed iterations. Its performance seems to be independent of the initial guess.

Table 1. Performance of the considered methods.

Initial guess	Method	$\frac{1}{k+1}$		$\frac{10}{k+1}$		$\frac{100}{k+1}$	
		Time	Iter	Time	Iter	Time	Iter
$x^0 = x^a$	OSGA	0.088	32	0.268	100	0.696	316
	ESGA	4.246	2228	4.357	2313	6.164	3288
	EPSM	0.018	1000	0.009	131	0.024	1279
$x^0 = x^b$	OSGA	0.089	32	0.273	100	0.689	316
	ESGA	2.495	1299	2.680	1400	4.730	2511
	EPSM	0.013	999	0.021	2087	0.046	4581
$x^0 = x^c$	OSGA	0.085	32	0.269	100	0.681	316
	ESGA	0.117	45	0.562	257	3.011	1572
	EPSM	0.011	999	0.068	8470	0.044	4999

4.2. Experiment 2

The key feature of the previous example is the paramononicity of the involved bifunctions. Requiring that property, the performance of ergodic version of Algorithm 1 is not comparative to the original version, and also to the method proposed in [27]. In order to highlight the advantage of the ergodic version of Algorithm 1, in this subsection we consider an example in which the involved bifunction is not necessarily paramonotone. For this example, both OSGA and EPSM methods cannot be applied due to the lack of the paramonotonicity of the involved bifunction, but ESGA can.

Let us consider the differentiated Cournot-Nash model similar to the previous example, but with linear cost functions

$$c_i(x_i) = \mu_i x_i + \xi_i \quad (\mu_i > 0, \xi_i \geq 0, i = 1, \ldots, n).$$

Furthermore, we assume that there are additional constraints concerning lower and upper bounds on quota of the commodity (*i.e.*, there exist $\underline{\sigma}, \overline{\sigma} \in \mathbb{R}_+$ such that $\underline{\sigma} \leq \sigma = \sum_{i=1}^n x_i \leq \overline{\sigma}$). The problem in this case can be formulated in form of (EP) in which we can split the bifunction f as $f(x,y) = f_1(x,y) + f_2(x,y)$ with

$$\begin{aligned} f_1(x,y) &= (Bx + \mu - \bar{\alpha})^T (y - x), \\ f_2(x,y) &= y^T \tilde{B} y - x^T \tilde{B} x, \end{aligned}$$

in which $\mu = (\mu_1, \ldots, \mu_n)$ and

$$B = \begin{bmatrix} 0 & \tau_{12} & \tau_{13} & \cdots & \tau_{1n} \\ \tau_{21} & 0 & \tau_{23} & \cdots & \tau_{2n} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \tau_{n1} & \tau_{n2} & \tau_{n3} & \cdots & 0 \end{bmatrix}, \quad \tilde{B} = \begin{bmatrix} \tau_{11} & 0 & 0 & \cdots & 0 \\ 0 & \tau_{22} & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & 0 & \cdots & \tau_{nn} \end{bmatrix}.$$

It is easy to check that f_1, f_2 are equilibrium functions satisfying conditions (A1)–(A3), so is f. By Theorem 3.4, this ensures that the ergodic sequence generated from the iterative points of Algorithm 1 converges without paramonotonicity of f. Therefore, in contrast to the previous experiment, in this experiment we do not need to check the paramonotonicity of the bifuction f.

For the tested instances, we set $C_i = [10, 50]$ for $i = 1, \ldots, n$, $\underline{\sigma} = 10n + 10$, $\overline{\sigma} = 50n - 10$, and $\alpha = 120$. The initial guess was set to $x_i^0 = 30$ $(i = 1, \ldots, n)$. The values for parameters τ_{ij} were randomly generated in $[0.5, 1.5]$ while the ones for μ_i were randomly chosen in $[20, 40]$.

Table 2 shows that the choice of parameter β_k is crucial for the convergence of the algorithm, since changing the value of this parameter may significantly reduce the number of iterations. Furthermore, it proves the applicability of the ergodic version of our proposed algorithm in solving differentiated Cournot-Nash model with linear costs and joint constraints.

TABLE 2. Performance of ergodic version of Algorithm 1 in solving differentiated Cournot-Nash model with linear costs and additional joint constraints.

n	β_k	Number of iterations
10	$10/(k+1)$	1309
10	$100/(k+1)$	419
15	$10/(k+1)$	3108
15	$100/(k+1)$	672
20	$10/(k+1)$	5090
20	$100/(k+1)$	610

4.3. Experiment 3

In this subsection we compare the performance of ergodic version of Algorithm 1 to the algorithm proposed in [2]. For that we consider the following instance of equilibrium problem (EP).

$$\text{Find } x \in C : f(x,y) \geq 0 \quad \forall y \in C,$$

in which $C = C_1 \times \ldots \times C_5$ with $C_i = [1, 10], i = 1, \ldots, 5$, and

$$f(x,y) = \langle Px + Qy - \bar{\alpha}, y - x \rangle + \sum_{i=1}^{5} \left(y_i^3 - x_i^3 \right),$$

where

$$P = \begin{bmatrix} 3 & 3 & 3 & 0 & 1 \\ 2 & 9 & 8 & 0 & 6 \\ 2 & 6 & 8 & 5 & 5 \\ 6 & 6 & 4 & 8 & 0 \\ 5 & 10 & 6 & 10 & 3 \end{bmatrix}, \quad Q = \begin{bmatrix} 21 & 23 & 17 & 15 & 21 \\ 23 & 50 & 36 & 21 & 18 \\ 17 & 36 & 76 & 27 & 60 \\ 15 & 21 & 27 & 25 & 27 \\ 21 & 18 & 60 & 27 & 66 \end{bmatrix}.$$

By setting

$$f_1(x,y) = \langle Px + Qy - \bar{\alpha}, y - x \rangle, \quad f_2(x,y) = \sum_{i=1}^{5} \left(y_i^3 - x_i^3 \right)$$

we see that $f(x,y) = f_1(x,y) + f_2(x,y)$. Note that $f_1(x,y)$ is quadratic but not separable with respect to y, so the subproblem of computing y^k in Algorithm 1 is nothing but solving a quadratic strongly convex program. Furthermore, $f_2(x,y)$ is not quadratic but separable with respect to y, so computing x^{k+1} in Algorithm 1 can be done by solving $n = 5$ separated convex programs, each on one real variable.

Table 3 presents the performance of ergodic version of Algorithm 1 (denoted ESGA) and the ergodic algorithm proposed in [2] (abbreviated by EA). For the former algorithm, we set $\beta_k = \frac{10^i}{k+1}$, while for the parameters in the description of the latter algorithm we choose $\lambda_k = \frac{10^i}{k+1}$ $(i \in \{0, 1, 2, 3\})$. We compare the performance of these algorithms in sense of running time. Three choices for starting points were taken:

$$x^a = (1, 10, 1, 10, 1), \quad x^b = (5.5, 5.5, 5.5, 5.5, 5.5), \quad x^c = (10, 10, 10, 10, 10).$$

It can be observed from Table 3 that, regardless of the choices for parameters β_k and λ_k, our proposed ergodic algorithm performs better than the one proposed in [2]. The reasons are as follows. By splitting the equilibrium function $f(x,y)$ as $f_1(x,y) + f_2(x,y)$, we then can exploit special properties of component bifunctions f_1 and f_2. More precisely, that helps us to solve more easily the two subproblems in each step of our proposed algorithm: one is a quadratic strongly convex program, while the other is equivalent to solving separated convex programs

Table 3. Performance of the considered ergodic algorithms.

Initial guess	Method	Running time (s)			
		$\beta_k = \lambda_k = \frac{1}{k+1}$	$\beta_k = \lambda_k = \frac{10}{k+1}$	$\beta_k = \lambda_k = \frac{100}{k+1}$	$\beta_k = \lambda_k = \frac{1000}{k+1}$
$x^0 = x^a$	ESGA	1.116	1.121	1.256	1.164
	EA	3.601	3.866	4.099	3.533
$x^0 = x^b$	ESGA	0.721	0.758	1.023	1.346
	EA	3.154	3.029	2.718	2.388
$x^0 = x^c$	ESGA	0.016	0.158	0.591	1.473
	EA	6.075	6.487	6.125	4.805

on single real variable. However, the sum $f = f_1 + f_2$ does not inherit the properties of the summand functions f_1 and f_2. The algorithm proposed in [2] applies directly on f and does not exploit special properties of f_1 and f_2. That leads to the worse performance of this algorithm in comparison to our proposed one.

5. Conclusions

We have proposed splitting algorithms for equilibrium problems where the bifunction is the sum of the two ones f_1 and f_2. At each iteration the proposed algorithms require solving two strongly convex programs, one for each f_1 and f_2 separately. Under a paramononicity property, the convergence of the iterates to a solution without any Lipschitz type condition as well as Hölder continuity of the bifunctions involved has been proved. It also has been shown that the ergodic sequence defined by the iterates of the algorithm converges to a solution without paaramonotonicity. We have applied the proposed algorithm to solve some versions of differentiated Cournot-Nash equilibrium model. Some numerical results have been reported for these models with different data. We have compared our algorithm to the one in [27] for paramonotone bifunction, and to the algorithm in [2] for monotone case. The obtained computational results have shown that for paramonotone problems the algorithm in [27] worked better in computing time, while for monotone problems the ergodic version of our algorithm runs more quickly than that in [2].

Acknowledgements. The authors would like to thank anonymous reviewers for their insightful comments and suggestions, that greatly improved this work. We would like to express our appreciation to Professor Le Dung Muu for his helpful comments and encouragement. The author X.T. Le is supported by Project ĐLTE00.02/19-20 of Vietnam Academy of Science and Technology.

References

[1] P.K. Anh and T.N. Hai, Splitting extragradient-like algorithms for strongly pseudomonotone equilibrium problems. *Numer. Algorithms* **76** (2017) 67–91.

[2] P.N. Anh, T.N. Hai and P.M. Tuan, On ergodic algorithms for equilibrium problems. *J. Global Optim.* **64** (2016) 179–195.

[3] H.H. Bauschke and P.H. Combettes, Convex Analysis and Monotone Operator in Hilbert Spaces. Springer, New York, NY (2010).

[4] G. Bigi, M. Castellani, M. Pappalardo and M. Passacantando, Existence and solution methods for equilibria. *Eur. J. Oper. Res.* **227** (2013) 1–11.

[5] G. Bigi, M. Castellani, M. Pappalardo and M. Passacantando, Nonlinear Programming Techniques for Equilibria. Springer, New York, NY (2018).

[6] E. Blum and W. Oettli, From optimization and variational inequalities to equilibrium problems. *Math. Stud.* **63** (1994) 123–145.

[7] P.M. Duc, L.D. Muu and N.V. Quy, Solution-existence and algorithms with their convergence rate for strongly pseudomonotone equilibrium problems. *Pac. J. Optim.* **12** (2016) 833–845.

[8] F. Facchinei and J.-S. Pang, Finite-Dimensional Variational Inequalities and Complementarity Problems. Springer New York, NY (2003).

[9] K. Fan, A minimax inequality and applications, edited by O. Shisha. In: Inequalities III. Academic Press, New York, NY (1972) 103–113.

[10] T.N. Hai and N.T. Vinh, Two new splitting algorithms for equilibrium problems. *Rev. R. Acad. Cienc. Exactas Fis. Nat. Ser. A Mat.* **111** (2017) 1051–1069.

[11] P.G. Hung and L.D. Muu, The Tikhonov regularization extended to equilibrium problems involving pseudomonotone bifunctions. *Nonlinear Anal. Theory Methods App.* **74** (2011) 6121–6129.

[12] A.N. Iusem, On some properties of paramonotone operators. *Convex Anal.* **5** (1998) 269–278.

[13] A.N. Iusem and W. Sosa, Iterative algorithms for equilibrium problems. *Optimization* **52** (2003) 301–316.

[14] I.V. Konnov and S. Schaible, Duality for equilibrium problems under generalized monotonicity. *J. Optim. Theory App.* **104** (2000) 395–408.

[15] G.M. Korpelevich, An extragradient method for finding saddle points and for other problems. *Ekon. Mat. Metody* **12** (1976) 747–756.

[16] G. Mastroeni, Gap functions for equilibrium problems. *J. Global Optim.* **27** (2003) 411–426.

[17] G. Mastroeni, On auxiliary principle for equilibrium problems, edited by P. Daniele, F. Giannessi and A. Maugeri. In: Vol. 68 of *Equilibrium Problems and Variational Models, Nonconvex Optimization and its Applications*. Springer, Boston, MA (2003) 289–298.

[18] A. Moudafi, On the convergence of splitting proximal methods for equilibrium problems in Hilbert spaces. *J. Math. Anal. App.* **359** (2009) 508–513.

[19] L.D. Muu and W. Oettli, Convergence of an adaptive penalty scheme for finding constrained equilibria. *Nonlinear Anal.* **18** (1992) 1159–1166.

[20] L.D. Muu and T.D. Quoc, Regularization algorithms for solving monotone Ky Fan inequalities with application to a Nash-Cournot equilibrium model. *J. Optim. Theory App.* **142** (2009) 185–204.

[21] L.D. Muu and N.V. Quy, On existence and solution methods for strongly pseudomonotone equilibrium problems. *Vietnam J. Math.* **43** (2015) 229–238.

[22] P.N. Natarajan, Classical Summability Theory. Springer, Singapore (2017).

[23] H. Nikaidô and K. Isoda, Note on noncooperative convex games. *Pac. J. Math.* **5** (1955) 807–815.

[24] G.B. Passty, Ergodic convergence to a zero of the sum of monotone operators in Hilbert space. *J. Math. Anal. App.* **72** (1979) 383–390.

[25] T.D. Quoc, L.D. Muu and N.V. Hien, Extragradient algorithms extended to equilibrium problems. *Optimization* **57** (2008) 749–776.

[26] T.D. Quoc, P.N. Anh and L.D. Muu, Dual extragradient algorithms extended to equilibrium problems. *J. Global Optim.* **52** (2012) 139–159.

[27] P.S.M. Santos and S. Scheimberg, An inexact subgradient algorithm for equilibrium problems. *Comput. Appl. Math.* **30** (2011) 91–107.

[28] B.F. Svaiter, On weak convergence of the Douglas–Rachford method. *SIAM J. Control Optim.* **49** (2011) 280–287.

[29] K.-K. Tan and H.-K. Xu, Approximating fixed points of nonexpansive mappings by the Ishikawa iteration process. *J. Math. Anal. App.* **178** (1993) 301–308.

[30] P.T. Vuong, J.-J. Strodiot and V.H. Nguyen, On extragradient-viscosity methods for solving equilibrium and fixed point problems in a Hilbert space. *Optimization* **64** (2015) 429–451.

[31] H.K. Xu, Iterative algorithms for nonlinear operators. *J. London Math. Soc.* **66** (2002) 240–256.